United Nations ECE/TRADE/C/CEFACT/2018/8

Economic and Social Council

Distr.: General
20 April 2018

English only

Economic Commission for Europe

UNECE Executive Committee

Centre for Trade Facilitation and Electronic Business
Twenty-fourth session
Geneva, 30 April and 1 May 2018
Item 7b of the provisional agenda
Recommendations and standards: Other deliverables for noting

White Paper on Data Pipeline Concept for Improving Data Quality in the Supply Chain

Summary

This White Paper builds upon the work that was undertaken by several European Union projects (Cassandra, CORE, SELIS) and which was brought to UN/CEFACT in order to provide a consistent definition and understanding of what is a Data Pipeline and how it should function. This paper also explores the possible information which can be input or output from a Pipeline at various "waypoints" along the life cycle of a transaction.

This document is submitted to the twenty-fourth session of the UN/CEFACT Plenary for noting.

GE.18-06366(E)

1806366

Please recycle

I. Background

1. The supply chain of today is much more complex than it has ever been, yet the data flows are still built around traditional paper-based processes that no longer satisfy today's requirements. The need for accurate, timely data through a supply chain is now a standard requirement from the commercial sector, yet the data provided by multiple parties is often re-keyed several times, typically open to interpretation or watering down to populate various types of documents that are passed between parties to support regulatory, financial or operational requirements. This is often done without consideration of where those documents will go or who will use them and for what purpose beyond the initial recipient.

2. This document-based approach does not adequately support today's requirements. With the increased automation that businesses are implementing into their warehouses and the advancements made in artificial intelligence (AI) and predictive analytics, the quality, timing and accuracy of data is more important than ever.

3. Consider the damage that inaccurate data could cause to a business when used unknowingly in decision making and planning, as well as financial forecasting. A white paper by Professors Alan Braithwaite and Richard Wilding estimates that data inaccuracy among the United Kingdom top five retailers and their suppliers is costing as much as 1.4 billion pounds sterling per year – 1 per cent of their total revenue.[1]

4. When commercial parties understand the risks, manage them, purchase goods using appropriate International Commercial Terms (Incoterms) and have visibility of their supply chain movements, they can manage the exceptions from real time data, and control the flow of goods into their business. This allows them to keep unexpected costs to a minimum and give a true and consistent landed cost for their goods. Businesses have demonstrated a 30 per cent reduction in their international supply chain costs using this approach.[2]

5. Border agencies also face increasing volumes of cargo traffic at borders, due to larger vessels/aircraft and an increasingly globalized approach to the way products are bought through less traditional methods, but now more commonplace channels, such as e-commerce platforms. Dealing with this increase in volume demands resources and the use of the correct data to enable efficient risk assessment and clearance to take place. It is indicated that a 50 per cent reduction in regulatory administration at the border could be worth 370 million pounds sterling.

6. However, poor quality data is all too often presented to these border agencies through the use of a manifest. This data is sometimes third or fourth hand and rarely comes from its true source. Freight forwarders often identify themselves or their overseas agent as the shipper and consignee on transport documents such as the bill of lading, making it increasingly difficult to identify the true parties related to the goods. In addition, industries like insurance companies may stipulate that 'terms' such as 'said to contain' need to be added to a bill of lading which leads to the use of data too imprecise when provided to the applicable border agency. Inaccurate or late data presents challenges and creates inefficiencies at borders around the world.

7. Trust in the data quality and source of the data is key for the industry as a whole to move away from the 'paper safety net'. Lack of trust in the data is hindering global trade facilitation and the move to true digitalization of the supply chain. Paper free initiatives in all modes of transport are yet to become widely adopted. For example, in sea freight many

1 SupplyChainStandard.com - Data inaccuracy costs retail 1.4 billion pounds sterling - viewed 2 July 2010: www.supplychainstandard.com/ICArticleID/29/Articles/9048/Data+inaccuracy+costs+retail+14+bn.html

2 D10.4 CORE Final report on phase two developments of the Felixstowe demonstrator

bills of lading are still made up of three originals and require six copies and even digitizing of the documents isn't acceptable for many.

II. Pipeline

8. The concept of a data pipeline allows data that originates at its source to be provided once and used multiple times throughout the supply chain, regardless of the mode of transport, party or border agency that needs to access the data.

9. In the pipeline concept, information is structured around waypoints when information is injected into the pipeline ("input waypoints") or when information is extracted from the pipeline ("output waypoints"). Various events take place in the process and movement of goods and these can potentially be "input waypoints", "output waypoints" or both.

10. There is commonality of these waypoints across all modes of transport. For example, goods need to be loaded onto a means of transport and need to be delivered to a location. Throughout the movement there are many waypoints that can take place which may trigger an important related event such as payment for the goods or transportation service which may demonstrate change of ownership or responsibility. These events can trigger a push of data or a pull of data for those that have the required access to this type of data.

11. During the Buy, Ship, Pay process that takes place, various data is captured and logged for use in the sales contract, invoicing process and the physical movement of goods. Rather than documents being passed through the chain on request or as a matter of course, the focus should be around making the required data available to those who need it for a specific purpose. This data should be captured upfront and made available using channels where the security of the data is contained and restricted for those that require it. This data should be entered into the pipeline as close to the original source as possible, and preferably directly from the systems that generate the data to show compliance and remove the possibility of illegal or erroneous tampering. The consignor knows more about the goods than any other stakeholder; hence, they are a key actor in the pipeline concept.

A. Principles

1. Capture the data from the right person at the right place at the right time

12. Capturing data at or as close to its source as possible is a key pipeline data carrier principle and offers multiple benefits to the actors in the supply chain. Data quality can be improved as the actors providing the data (the right person) access it at its origin (the right place) and so are able to provide reliable and detailed data elements to the Pipeline Data Exchange Structure (PDES).

13. Capturing key information at its source, rather than later on in the supply chain, removes the need to rely on data which has been re-entered by other actors, eliminating the risk of error or data misinterpretation. Data availability is also enhanced (the right time), as data will be input earlier and hence available for inclusion in data sets destined for cross-border agencies. Logistics and commercial planning and organization can also benefit.

14. For some, capturing the data from the source can be a challenging situation. For many actors in the supply chain they may resist as they have never had to provide this type of data before. Commercial understanding of the reasons why this is important need to be explained. It could form part of the sales contract or price negotiation as it would be considered a valued service and savings would be obtained further down the supply chain.

15. Regulatory bodies are increasingly upstreaming the compliance frontier further up the supply chain, for example in air freight the Pre-Load Advance Cargo Information for security filing is demanding data earlier in the transportation movement.

2. Capture once and use many times in the supply chain

16. A second important principle is that not only is pipeline data collected early on in the movement of goods along the supply chain, but once a data element is included in the pipeline data exchange structure, it does not need to be re-entered. This reflects the single window principle of sharing electronic data as defined in UNECE Recommendation 33.[1] Data re-entry is reserved exclusively for data elements that have changed or need correcting. The pipeline data exchange structure makes it possible for the data elements to be used by multiple cross-border agencies, avoiding the need to resubmit data for each agency. For certain types of data such as parties (buyer, seller) the use of external trusted data source(s) could be included to provide certainty.

B. Capturing information at source

17. Data captured from the source of the information provides more certainty that the information provided is correct at that point in the process. The consignor is a key actor to start the data capture in the transport process. The consignor knows more about the goods than any other actor in the supply chain at that point in time and in most cases, they are the last actor to physically touch the goods before being sealed for transport. This data used for customs declarations (export or import) will be more faithful to reality as declarants can provide information coming from the source.

18. The transport provider will also provide key insight around the journey of the shipment. Many road hauliers provide GPS tracking data. Ports, sea freight and air freight carriers, provide actual events that take place in the transport process, that are beneficial commercially as well as from a risk perspective. For example, ports of call on a shipment, arrival or departure dates all mark key points and are available electronically.

19. More advanced providers of data exist and can easily be integrated into the pipeline using more digital automated methods. For example, "Internet Of Things" (IoT) devices on containers, at ports and weigh bridges all provide the opportunity to 'digitally link' the events or waypoints of transport movements.

20. Participating businesses which are prepared to voluntarily provide advanced data can and will also provide input data back to the pipeline. For example, 'goods receipt' at their warehouse and confirmation of package totals received. It is beneficial for them to do this as the data could be fed back through output waypoints to the customs authorities to correct or amend the import customs declaration from the simplified declaration made before the container is opened or pallet unpacked.

21. Sources of data must be augmented to provide the whole data set for the PDES. No single party can capture everything for every shipment as the nature of supply chains is that it is multi party. For this reason, data provided should be limited to that which is originating and known to be true to that source, at that point in time. The data around the goods, augmented with transport data and security data, forms a complete and neutral picture of what actually took place during the movement.

---

[1] UNECE Recommendation 33, 2005.
http://www.unece.org/fileadmin/DAM/cefact/recommendations/rec33/rec33_trd352e.pdf

C. Output waypoints: extracting information from the pipeline

22. Data contained in the pipeline can be extracted by the authorized actors by using two methods, push or pull.

- Push Methodology - Some regulatory agencies desire data to be pushed to their data repositories. They then store this data and use it to assess risk and provide historical retrospective in their own internal systems.
- Pull Methodology - Some regulatory agencies desire a pull process where and when they can request data using a secure access token or similar. They plan to evaluate risk internally in their own systems. If a shipment is selected for inspection they can request the data and its history to re-evaluate the risk and make a decision without the need for a physical inspection or manual intervention to provide documentation.

23. Both methods have their advantages and drawbacks; however, commonly both require the use of a structured, standardized data model. This reduces the development effort and the data analytics energy required to make sense of the data and provide certainty of the data fields of interest to the local regulatory agencies.

24. This push and pull methodology can be adapted for output waypoints between commercial actors as well. In these cases, the relevant privacy and data ownership aspects would need to be addressed bilaterally or multilaterally. These methods could also be used for data exchange between data pipeline providers.

III. Building a pipeline carrier

A. Basic principle

25. It is important to identify the actors who will be providing the input waypoint data ahead of time and enable them to contribute to the PDES.

26. When building a PDES there needs to be a clear path for inbound data to be captured at the right place and added to the overall pipeline carrier. The data would be added in blocks or snippets which relate specifically to new or updated data that is being captured from that source. It is important that data that is specific and known to be true at the input waypoint is what is being completed, and not to provide 'passed on' or 'hearsay' information that cannot be proven. Some input waypoints will contain 'planned' data; for example, the route plan for the transport movement. This type of data can be used to predict or estimate when something should be happening and provide exceptions when it does not happen. This, in turn, improves business planning, targeting and risk assessment.

27. The use of web services would assist with the capture and validation of the information for the providers and recipients, given that the stakeholders agree on message protocols and related services.

28. Ideally, the captured data should be timestamped and recorded for auditing, as well as identifying the source and location of the data capture for further analysis by risk assessment tools.

29. Missing data can be identified by exception management. The transport departure confirmation would be expected by the planned time noted in an earlier input waypoint. If it is missing, the data can be queried with the relevant party or source automatically.

30. Due to the number of actors and data capture points that could be automated using IoT devices, the environment is well suited to the use of Distributed Ledger Technology (DLT) such as Blockchain.

B. Use of standards

31. The PDES is intended to be a model that anyone could use and eventually exchange from one Pipeline data carrier to another. If this is to be the case, the use of internationally recognized standards is essential. Information must be unambiguously understood between the sender and any party that is authorized to receive the information.

32. The standard is important on several levels, including:

i) Firstly, the data element level to ensure that the semantics are understood to be the same across all participants.

ii) It is recommended to codify information when possible in order to avoid misunderstandings due to language or spelling differences; such code lists and data types must also be standardized in order to be mutually understood.

iii) The message structure should also be standardized as information is very rarely flat, but it is often hierarchal bringing information into relation with other information (for example, a seal should be associated to the container on which it is affixed; or the postal code should be associated to the party to which it is related).

iv) The standardization of the exchange protocol can also facilitate the exchange of information, though this can be less critical than the three other levels of standards.

33. The UN/CEFACT Pipeline project includes this paper and the technical artefacts that standardize the data, the code lists, the data types, the message structure (the PDES) and it also provides a suggested XML message format for the data exchange. UN/CEFACT standards are the most pertinent for this type of message because of the inherent business nature of the information being shared and the potential business and regulatory use of these. All of these deliverables are available on United Nations websites free of charge and all registered experts are invited to participate in its development and subsequent maintenance, also free of charge.

C. Input waypoints: adding data into the carrier

34. Information can be integrated into the PDES as a submission of a full document. However, in this case, information may be redundant as documents often repeat information that was established earlier in the data exchange. In this case, redundant information should not overwrite previously submitted information unless it could be considered the data source (for example, the estimated time of departure of the means of transport might be provided earlier in the transaction, but the operator of the means of transport is the source for the actual time of departure and their data should be able to overwrite a previously sent estimation).

35. Information can be integrated as part of an exchange of data in a process-oriented approach. An Enterprise Resource Planning tool (ERP) might be able to provide snippets of information as they become available instead of waiting for the establishment of a traditional document. Similarly, information can be integrated through IoT devices such as smart containers, smart devices, warehouse scanners or other.

D. Change of information

36. Data captured through the pipeline should in almost all cases be actual data that is true at that point in time. The amending of data should not as a rule take place; however, some waypoints can capture and provide the 'planned' route or 'estimated' arrival dates, which by nature are variable. At each input waypoint, the latest advised estimates for these events or factors could be provided.

37. If information is found to be incorrect after the time that it was provided it could be resent and could update the pipeline carrier; however, the new data should be time stamped and a flag set to indicate change after the event to provide those looking at pipeline data assurance that the data hasn't been tampered with. Only the original actor should be able to amend their waypoint data to maintain integrity. The challenge that amending data provides is one of lack of trust. If the data changes frequently or is being 'cleaned' before sending, it will not enable true automation in the supply chain. For example, if the data is being used in a distributed ledger technology, a 'smart contract' may be triggered once the data has entered the ledger. It is then too late to 'roll back' the outcome of that contract. For this reason, the changing of information once submitted should be carefully considered if appropriate for the intended use case.

IV. Data waypoints

38. Waypoints are the definitive moments in a supply chain when information is exchanged. Not all waypoints need to be used each time, but there should be a defined list so that pipelines can be implemented across all modes of transport and have a common understanding by all.

39. Some common input waypoints could be:

Event / Input Waypoint Status Code from UNECE Rec24	Type of Data	Actor / Role / Source	Description
Sales contract	• Buyer • Seller • Despatch and delivery party • Country of origin • Harmonized System Codes • Commercial document reference (invoice number or other) • Commercial (invoice) value of goods • Quantity of goods ordered • Planned place of delivery • Payment and delivery conditions such as Incoterms	Buyer / seller	The sales contract is the source of many of the subsequent exchanges on a supply chain. It provides details about the goods that are being sold (commercial value, HS codes…) and the parties that are involved. An important part of risk analysis and logistics planning for the PDES finds its source in this event.

Goods booked for transport Rec24 status code 6 (Booking Completed)	• Number of packages for shipment • Weights • Special handling instructions (perishables, dangerous goods information, temperature control) • Means of transport ID • Port of loading • Planned time of departure • Planned port of discharge • Planned transport route • Freight rate	Transport service buyer or transport service provider	The point at which goods are booked with a transport service provider and space allocated for that movement. The transport may take place over multiple shipments under a single transport-sales contract, so some of the information provided from previous waypoints would need to be provided again for that specific movement, for example number of packages.
Transport contract (shipping instruction)	• Actual point of departure/collection • Actual date of collection • Carrier • ETA - estimated time of arrival	Transport service buyer or transport service provider	The instruction for shipment and production of the transport contract (e.g. bill of lading or road consignment note CMR) from the transport service provider to the carrier. Usually done before the shipment takes place but not unusual to provide after the shipment has taken place.
Goods loaded for transport Rec24 status code 13 (Collection / Pickup Completed)	• Actual arrival time at pickup place / the place of acceptance • Actual departure time from pickup place / the place of acceptance • Container number • Seal ID(s) • Container verified gross mass (VGM) weight [method 2]	Transport service provider	The point that the goods are physically loaded onto or into a means of transport to begin the transportation. The goods may be 'sealed' into a loading device such as a container or unit load device (ULD). This is the point when liability for the goods is taken over by the Transport Service Provider under the terms of the Transport Contract.
Arrived at port of loading Rec24 status code 364 (Arrival at Yard)	• Actual time of arrival at the port of loading • Power or plug in provided for temperature-controlled goods and devices • Container verified gross mass (VGM) weight [method 1]	Transport service provider	The point at which goods arrive at the port and change of liability usually takes place from the transport service provider to the port or carrier.
Goods Departed the Port of Loading	• Actual time of departure • Vessel or Flight name/number • Actual Departure Date • Updated ETA	Transport service provider	The point at which the goods have actually left the port of loading. Confirming the transport has taken place.
Goods sold whilst in transport Rec24 status code 75 (Consignee	• Commercial document reference (invoice or other) • New buyer • New seller • Planned place of delivery	Buyer/seller	Some goods are sold whilst in transit, this can occur multiple times for some commodities. A new commercial transaction is established in this case which would supplement the original commercial contract information.

Changed)			
Transhipment (Relay) or transit Rec24 status code 100 (Transshipment)	• Means of transportation ID • Vessel or flight number • Time of transhipment • Location of transhipment • Updated ETA	Transport service provider	Transhipment is the transfer of the goods to another vessel while in transit. A transit point for air is an intermediate stop through which the goods pass and continue on toward final destination. The consignment may be unloaded and loaded onto another aircraft or just stay on the same conveyance. Generally, the consignment remains on the same mode of transport with the same transport service provider. For some modes of transport this could lead to a 'split consignment' where the goods may be separated into partial consignments and moved on a different conveyance number.
Transfer	• New carrier • Means of transportation ID • Vessel or flight name or number • Time of vessel change • Location of vessel change • Updated ETA	Transport service provider	The point at which the consignment changes onto another vehicle/vessel/aircraft, usually on the same mode of transport, but liability changes to a different transport service provider. For some modes of transport this could lead to a 'split consignment' where a partial consignment is transferred and moved on a different transport service provider's conveyance.
Import clearance procedure, including temporary storage and transit Rec24 status code 128 (Awaiting Import Release)	• Intended usage of the goods (commercial or not) • Customs value • Customs destination (direct import, storage, transit, special procedure) • Customs origin (according to importing-national legislation) • Authorization numbers (such as storage facility authorization number) • Authentication (of the declarant taking legal responsibility of the information)	Declarant	The point where the required customs declaration for import or other cross-border regulatory procedure takes place. Much of the information necessary for this output waypoint will be originating from previous input waypoints, however some additional information will probably be required.
Goods arrived at the final port of discharge	• Actual route taken • Actual arrival time	Transport service provider	The point at which the consignment physically arrives at the port of discharge which are then under customs supervision and possible

Rec24 status code 40 (Arrival at Port)			control.
Customs status Rec24 status codes 12 (Cleared by Customs), 17 (Customs Clearance Refused) and 379 (Cleared for Release)	• Regulatory decision (goods cleared, released, or on hold) • Time of regulatory decision	Customs or border agency	Customs or other regulatory agencies review the shipment for possible violations and value determination. A broker works to make sure the "paperwork" or ideally advanced data is in order and if passed by all required agencies, the goods are considered cleared. This could in future be released prior to arrival if all checks are in place.
Goods departed the port of discharge Rec24 status code 365 (Gate Out)	• Actual departure time	Transport service provider	The point at which the goods have actually left the port of discharge, on another means of transport other than that on which they arrived. Depending on the Incoterm, this could also be the delivery point where the consignment changes responsibility to the consignee.
Goods delivered Rec24 status code 21 (Delivery Completed)	• Actual delivery point • Actual date/time of delivery • GPS position of delivery point • Acceptance of delivery (person) / goods receipt	Consignee or transport service provider	The point at which the consignment is actually delivered to the consignee or dispatch party under the transport contract.
Doors opened Rec24 status code 309 (Equipment Opened)	• GPS position (versus the trip plan)	Customs, border agency or delivery party	Doors opened by customs, border agency or final delivery party. The seal would be broken.
IoT alert	Alerts whenever the physical parameters change outside of set parameters • Doors opening • Shocks and vibration • Temperature • Humidity • Pressure • GPS position for geo-fence trigger	IoT or software provider	Alerts could be provided for a number of reasons such as temperature out of range, goods dropped or within a geo-fenced area.

40. Waypoints and their use for input or output could be interchangeable and the use would be defined in the specific implementation, as would the data have extracted for the required purpose. Some classic output waypoints for example could be:

Event / output waypoint	Example Type of Data	Actor / Role / Source	Description
Door closing	• Container number • Seal number • Actual date of loading • Actual place of loading • No packages • Weight / cube	Consignor or transport service provider	The point that the goods are physically loaded onto or into the main means of transport to begin the transportation. The goods may be "sealed" into a loading device such as a container or ULD.
Export declaration	• Existing export procedure data requirements (UN/EDIFACT CUSDEC, customs declaration message)	Declarant	The point at which the required customs declaration for export takes place.
After departure	• Actual country of export • Actual departure date • Carrier (if different to planned)	Transport service provider	The point at which the goods have actually left the port of loading. Confirming the transport has taken place.
Just before arrival	• ETA update • Actual arrival location (if different to schedule) • Actual ports of call (itinerary) • Port of entry (i.e. UN/LOCODE)	Transport service provider	The point at which the main means of transport is due to arrive at its intended destination or port. Could be triggered by geo-fence for IoT enabled devices, or from Automatic Identification System (AIS) data for marine transport, for example.
Import declaration	• Intended usage of the goods (commercial or not) • Existing import procedure data requirements (UN/EDIFACT CUSDEC)	Declarant	The point where the required customs declaration for import takes place.
Goods delivered	• Actual delivery point • Actual date/time of delivery • GPS position of delivery point • Acceptance of delivery (person)	Consignee or transport service provider	The point at which the shipment is actually delivered to the consignee or dispatch party under the contract.